ISPRS Journal of Photogrammetry and Remote Sensing 137 (2018) 84–96

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Influence of conversion on the location of points and lines: The change of location entropy and the probability of a vector point inside the converted grid point

Nan Chen

Spatial Information Research Center, Fuzhou University, Floor 8, Science Park, No. 2 Xueyuan Road, Fuzhou 350116, Fujian Province, China
Key Laboratory of Spatial Data Mining & Information Sharing, Ministry of Education, Fuzhou University, Floor 8, Science Park, No. 2 Xueyuan Road, Fuzhou 350116, Fujian Province, China
Department of Geography, University of South Carolina, Callcott Building, 709 Bull Street, Columbia 29028, SC, USA

ARTICLE INFO

Article history:
Received 5 June 2016
Received in revised form 15 January 2018
Accepted 15 January 2018

Keywords:
Conversion
Probability theory
Entropy
Grid data model
Vector data model

ABSTRACT

Conversion of points or lines from vector to grid format, or vice versa, is the first operation required for most spatial analysis. Conversion, however, usually causes the location of points or lines to change, which influences the reliability of the results of spatial analysis or even results in analysis errors. The purpose of this paper is to evaluate the change of the location of points and lines during conversion using the concepts of probability and entropy. This paper shows that when a vector point is converted to a grid point, the vector point may be outside or inside the grid point. This paper deduces a formula for computing the probability that the vector point is inside the grid point. It was found that the probability increased with the side length of the grid and with the variances of the coordinates of the vector point. In addition, the location entropy of points and lines are defined in this paper. Formulae for computing the change of the location entropy during conversion are deduced. The probability mentioned above and the change of location entropy may be used to evaluate the location reliability of points and lines in Geographic Information Systems and may be used to choose an appropriate range of the side length of grids before conversion. The results of this study may help scientists and users to avoid mistakes caused by the change of location during conversion as well as in spatial decision and analysis.

1. Introduction

Conversion of spatial data from vector to grid format, and vice versa, is a common and important operation of Geographic Information Systems (GIS) and the first step for most spatial analysis. Conversion is concerned with the algorithms of conversion and the data before and after conversion. The algorithms have been evaluated and improved extensively (Piwowar et al., 1990; Schmidt and Tsetskhladze, 2013; Wang et al., 2013), thus getting mature; by comparing the data before and after conversion, researchers have found that conversion leads to conversion errors (Knaap, 1992; Wade et al., 2003), as map operations do (Haining and Arbia, 1993).

The studies on the conversion error have gone through two stages. In the first stage, scientists evaluated the conversion error, proposing a few methods for evaluation (Bregt et al., 1991), and found that the conversion error presented themselves in various different forms, including area errors (Liao and Bai, 2010), length errors and perimeter errors (Theobald, 2000), etc. In the second stage, a few specific models for reducing the conversion error were proposed, e.g., the equal area conversion model for reducing the area error (Zhou et al., 2007) and the model with correction coefficients for reducing the perimeter error (Theobald, 2000); meanwhile, to reduce the conversion error more effectively, scientists began to study the causes of the conversion error (Bettinger et al., 1996; Shortridge, 2004).

The causes for the conversion error, which are diversified, may be classified into two categories.

- The first category is related to the operation of conversion. The causes could be the cell assignment method of the operation (Diaz-Varela et al., 2010), the size of grids (or cells) chosen in the operation (Dendoncker et al., 2008; Diaz-Varela et al., 2010), and the unit of length (Wade et al., 2003) chosen in the

---

E-mail address: chennan@fzu.edu.cn

https://doi.org/10.1016/j.isprsjprs.2018.01.011

operation. The causes of the category could be reduced by choosing proper assignment method and appropriate parameters, e.g., an appropriate size of grids (Carver and Brunsdon, 1994).

- The second category is related to the source data of conversion (conversion is done to the source data). The causes of the category could be the complexity of the source data (the map) (Bregt et al., 1991), the polygonal structure and shape in the source data (Shortridge, 2004), the polygonal size in the source data (Yang et al., 2001), etc.

  In addition, the causes of the second category were studied when source data were divided into its geometrical elements—points, lines and polygons. Points and lines have the uncertainty of location (Goodchild et al., 1998; Shi, 2005) thus contributing to the conversion error (Goodchild and Hunter, 1997). Then, the causes of the category may include the uncertainty of the locations of the points and that of lines— we do not discuss the uncertainty of the location of polygons and its contribution to the conversion error in this paper, for the discussion may be quite complicated and be made in another paper.

The uncertainty of the locations of points (or lines) could be described by using three concepts of probability density function (PDF), error band and entropy— the corresponding discussion on the three concepts is as follows.

(1) The uncertainty of the locations of vector points could be described by using the concept of PDF. The PDF of a vector point $A(x_A, y_A)$ meets Bivariate Normal Distribution (Goodchild, 2004, 2008a), with the mean of $A(x_A, y_A)$ being $A_0(x_0, y_0)$. By the PDF of A, A is located around A_0 or at A_0 — A has the uncertainty of location before conversion.

The uncertainty of the location of A contributes to the conversion error of A. When A is converted to a grid point G, A_0 should lie inside G (according to the conversion algorithm) and there will be two cases for A—A will be inside G (see Fig. 1(a)) or outside G (see Fig. 1(b)). The second case may make a contribution to the conversion error, which will inevitably be propagated through spatial analysis (e.g., overlay analysis, query analysis.) on G. The probability of the first case, however, has not been computed before. Then we present Eq. (2) in this paper for computing the probability theoretically and Eq. (A.2) for its estimating in practice in this paper (the probability of the second case is computed by extracting that of the first case from one).

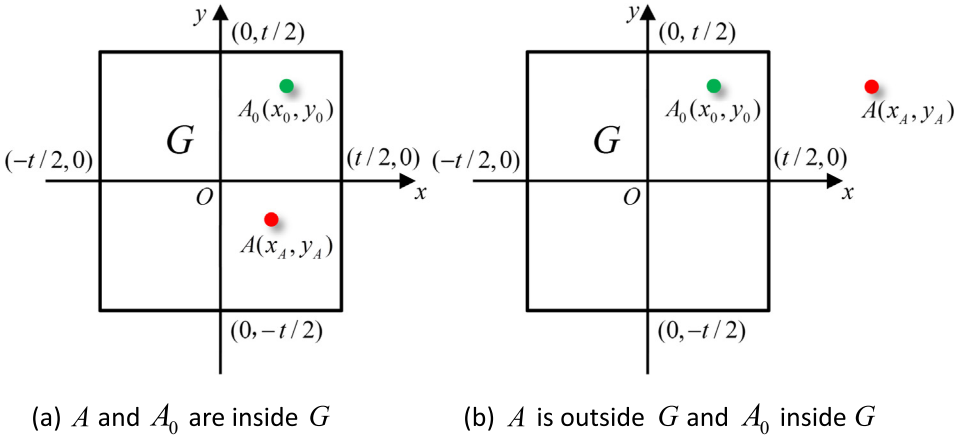
(2) In addition, the uncertainty of the locations of vector points and that of vector lines could be described by using the concept of error band. The location deference between A and "the real point" represented by A is the location error of A, which varies with direction; the varying error could be described by the error ellipse of $A(x_A, y_A)$ defined by the PDF of (x_A, y_A) with a confidence interval used to describe the probability that "the real point" is inside the error ellipse, by Surveying Adjustment Theory. Then by the PDFs of the points along a vector line, we could obtain the corresponding error ellipses of the points. Combining the error ellipses yields an error band—a band around the "real line" represented by the line (Perkal, 1956). The typical error bands of a vector line are as follows.

- The ε-band (Perkal, 1956), which expresses the isotropy of the errors of the points along the vector line and which has constant width. Then ε -band was improved because a method for determining the width of ε-band (Goodchild and Hunter, 1997) was proposed.
- The E-band (Caspary and Scheuring, 1993), whose width varies with the mean square error of the coordinates of the points along the vector line.
- The G-band model (Shi and Liu, 2000; Shi, 2005), which expresses the anisotropy of the errors of the points along the vector line.

The error bands mentioned, however, are dependent on their confidence intervals (Fan and Guo, 2001) and then the concept of entropy, which is independent on confidence intervals, has been used to describe the uncertainty of location.

(3) Entropy was used to describe the uncertainty of location error of vector points (Lin and Zhang, 2006). The location error (or the measurement error of location) of a vector point is an error vector quantity whose mean is the zero vector quantity (Leung et al., 2004a); the entropy of the Univariate Normal distribution was used to define the uncertainty of vector points (Lin and Zhang, 2006).

Meanwhile, on the basis of the ε-band, the H-Band model (Fan and Guo, 2001) was developed by using the concept of the maxi-



(a) A and A_0 are inside G (b) A is outside G and A_0 inside G

Fig. 1. Conversion of vector point, $A(x_A, y_A)$, to grid point, G. (x_A, y_A) are usually described by a two-dimensional random variable which satisfies the Bivariate Normal Distribution. Let the mean of $A(x_A, y_A)$ be $A_0(x_0, y_0)$. $A_0(x_0, y_0)$ should be inside G according to the algorithm for conversion. $A(x_A, y_A)$, however, may be inside or outside G.